

所要時間に対する認知集団の潜在クラスタリング

山口 正会員 ○岡本知也
 広島大学 正会員 藤原章正
 広島大学 学生員 李 百鎮

1. 本研究の背景及び目的

公共性の高い道路交通情報提供を行う際、どの程度詳しい所要時間で示せばよいか今のところ明確な結論は得られていない。その理由の 1 つとして、運転者の中に経路に対する所要時間の認知が異なる集団が混在することが考えられる。例えば、提供された時間を厳密に認知する集団や、あいまい性をもって認知する集団が存在する可能性がある。情報提供に伴う交通行動の変化を予測するためには、これらの異なる認知集団に対応する行動予測モデルの作成が望まれる。

そこで、本研究は母集団が上記の同質な 2 つの集団からなると仮定し、全被験者集団を 2 つの集団に分類することを目的とする。分析対象として高速道と一般道の 2 経路選択問題を取り扱い、集団の分類には混合分布を構成するいくつかの正規分布を判別することができる EM アルゴリズムを用いた潜在クラスタリング法を用いる。

経路選択モデルの構築にあたり、所要時間をあいまいに認知し、ファジィ理論に依拠した経路選択の意思決定プロセスを踏む集団をファジィ集団と定義する。一方、所要時間をはっきりと認知し、ランダム効用理論に依拠したプロセスを踏む集団をクリスプ集団と定義する。前者の経路選択行動はファジィ推論モデルにより、後者の経路選択行動はロジットモデルにより各々記述できるものとする。

2. EM アルゴリズムによる集団の分類方法

図 1 は、観測項目の分布の中にいくつかの正規分布をなす集団が混在する混合分布を表したものである。観測される混合分布の中から潜在する正規分布を見つけることが本分析のねらいである。

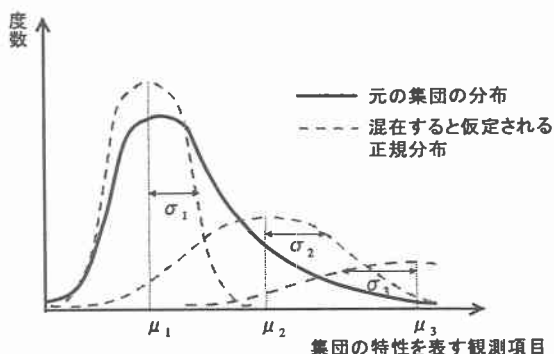


図 1：観測項目の分布に混在する正規分布をなす集団

標本数 n の観測データを $\mathbf{w} = (w_1, \dots, w_n)$ とする。このとき、 $\mathbf{w}$ には正規分布に従ういくつかの集団が存在すると仮定する。運転頻度のような各集団の特性を表す観測項目の平均値 μ_i と分散 σ_i^2 をとする ($\theta_i = (\mu_i, \sigma_i^2)$)。この際、 $i = 1, \dots, g$ は集団の数を表すとすると、推定パラメータは次式で定義できる。

$$\Psi = (\theta_1, \pi_1, \dots, \pi_{g-1}) \quad (1)$$

$$\pi_g = 1 - \sum_{i=1}^g \pi_i \quad (2)$$

ここで、正規分布の確率密度関数は、

$$\begin{aligned} f(w_i; \Psi) &= \sum_{i=1}^g \pi_i f_i(w_i; \theta_i) \\ &= \sum_{i=1}^g \pi_i \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{(w_j - \pi_i)^2}{2\sigma_i^2} \right] \end{aligned} \quad (3)$$

と表されるので、個人 j の集団 i への帰属を表す指標 $z_{ij} = (0, 1)$ を導入することにより、尤度関数は次式で表される。

$$L_c(\Psi) = \prod_{i=1}^g \prod_{j=1}^n \{ \pi_i f(w_j, \theta_i) \}^{z_{ij}} \quad (4)$$

このとき f は正規分布の確率密度関数を表す。

式(4)のパラメータの推定にあたり、アルゴリズムの単純性とパラメータ収束に対する安定性(収束が生じ易く時間がかからない性質)より、ここでは EM アルゴリズムを採用する。計算手順は次のとおりである。

①E-step：集団への帰属に関する条件付期待値の計算。

$$z_{ij}^{(k)} = \left\{ \pi_i^{(k)} f_i(w_j; \theta_i^{(k)}) \right\} / f(w_j; \Psi^{(k)}) = (0, 1) \quad (5)$$

π_i 及び $\theta_i = (\mu_i, \sigma_i^2)$ に初期値 π_1 及び $\theta_1 = (\mu_1, \sigma_1^2)$ を代入し、平均値及び分散を次式により求める。

$$\mu_i^{(k+1)} = \left(\sum_{j=1}^n z_{ij}^{(k)} w_j \right) / \sum_{j=1}^n z_{ij}^{(k)} \quad (6)$$

$$\sigma_i^{2(k+1)} = \left\{ \sum_{j=1}^n z_{ij}^{(k)} (w_j - \mu_i^{(k+1)})^2 \right\} / \sum_{j=1}^n z_{ij}^{(k)} \quad (7)$$

②M-step：これらを式(4)の確率密度関数 f に代入して求まる対数尤度の条件付期待値(8)を最大化する。

$$\pi_i^{(k+1)} = \left(\sum_{j=1}^n z_{ij}^{(k)} \right) / n \quad (8)$$

③E-step 及び M-step を π_i が収束するまで反復計算することで、最終的に帰属度 z_{ij} 及び Ψ が得られ、集団は分類される。

3. 同質な認知集団の分類結果

集団の分類を行うに当たり 2 つの仮定を設定する。

仮定 1: 高速道と一般道の総運転頻度 (Hex+Lex) が多ければ、よりクリスプ集団に近づく。

仮定 2: 各経路の運転頻度の差(|Hex-Lex|)が大きければ、よりファジィ集団に近づく。

そこで、Hex+Lex 及び|Hex-Lex|を観測項目(w)と考えて、EM アルゴリズムを用いて集団を 3 つ(AIC による最適分類数)に分類した。結果を表 1、表 2 に示す。

まず表 1 から、集団は Hex+Lex の基準より、10 回以上の集団 A、4~9 回の集団 B、3 回以下の集団 C の 3 集団に分類された。それぞれの混合分布の平均値、分散は表に示す通りである。また、経路選択率は分類後の集団それぞれに占める経路選択別割合を表す。

また、|Hex-Lex|の基準により、9 回以上の集団 D、3~8 回の集団 E、2 回以下の集団 F の 3 集団に分類された。経路選択、平均値、分散も表に示すように、集団により基準の平均値、分散、経路選択率がことなることが明らかである。

表 1: Hex+Lex による分類結果

集団	Hex+Lex (回/月)	出力平均値	出力分散	サンプル数	経路選択率	
					高速道路	一般道路
A	10~45	17.96	98.03	36	36%	64%
B	4~9	5.02	3.41	86	38%	62%
C	0~3	2.03	0.68	162	51%	49%

表 2: |Hex-Lex|による分類結果

集団	Hex+Lex (回/月)	出力平均値	出力分散	サンプル数	経路選択率	
					高速道路	一般道路
D	9~30	16.08	75.91	15	27%	73%
E	3~8	3.93	3.50	73	45%	55%
F	0~2	1.27	0.70	196	47%	53%

4. 分類精度の評価と経路選択モデルの推定

分類後各集団に、従来の行動予測モデルである 2 項ロジットモデル(BLM)及び、ファジィ推論モデル(FRM)を適用し、モデルの的中率により分類精度の比較を行う。ここでいう的中率とは各モデルにより再現される経路選択結果と、観測された経路選択が一致する割合を示したものである。

3 章の仮定に基づくと、Hex+Lex が多く|Hex-Lex|が小さいクリスプ集団は BLM の的中率が上がり、FRM の的中率が下がる。一方、Hex+Lex が少なく|Hex-Lex|が大きいファジィ集団は BLM の的中率が下がり FRM の的中率が上がるという結果が予測される。

Hex+Lex による分類の場合は表 3 の結果を得た。分類前全集団データの的中率を比較して、集団 A(総運転頻度 10 回以上)では BLH の的中率が高く、集団 C(総運転頻度 3 回以下)では FRM の的中率が高い、従って、集団 A はクリスプ集団、集団 C はファジィ集団と判別するこ

とができる。一方、集団 B(総運転頻度 4~9 回)は BLM、FRM 共に的中率が低いために、ファジィ集団からクリスプ集団への移行過程にあると解釈でき、どちらの集団に属するか判別が難しい。

|Hex-Lex|による分類の場合は表 4 の結果を得た。集団 D に関してサンプル数が少ないため(表 2 参照)、ここでは集団 D と集団 E を合わせた集団 D+E(経路運転頻度の差が 3 回以上)と集団 F(経路運転頻度の差が 2 回以下)の比較を行った。分類前全集団と比べて、集団 D+E は BLH の的中率が低く、FRM の値が高いためファジィ集団、集団 F は BLM の的中率が高く FRM の的中率が低いいためクリスプ集団と判別することができる。

表 3: Hex+Lex による分類集団の判別結果

集団	BLM	FRM	判別結果
分類前全集団	0.827 (μ)	0.859 (μ)	
C	0.895 ($>\mu$)	0.901 ($>\mu$)	ファジィ
A	0.944 ($>>\mu$)	0.860 ($\approx\mu$)	クリスプ
B	0.826 ($\approx\mu$)	0.813 ($<\mu$)	不可

表 4: |Hex-Lex|による分類集団の判別結果

集団	BLM	FRM	判別結果
分類前全集団	0.827 (μ)	0.859 (μ)	
D+E	0.821 ($<\mu$)	0.901 ($>\mu$)	ファジィ
F	0.842 ($>\mu$)	0.813 ($<\mu$)	クリスプ

5. 結論

本研究により、交通情報提供を行う際には、このような個人による認知の違いを考慮することは大変重要であることが再確認された。

EM アルゴリズムを用いた潜在クラスタリング法により、Hex+Lex 及び|Hex-Lex|の項目を基準として、全集団をファジィ集団とクリスプ集団の 2 つの集団に分類することができた。ただし、Hex+Lex を基準とした際、一部の集団が 2 つのいずれの集団にも分類されなかった理由は、分類基準項目を経路の運転頻度のみとしたため、分類基準の説明力の不足によるものと考えられる。

EM アルゴリズムにより導出された分類集団への帰属を表す指標 z_y は、今後、ランダム効用理論とファジィ理論を統合した行動予測モデルの構築の際に有力なツールとなるであろう。

参考文献

- 1) McLachlan,G.J. and Krishnan,T. (1997): The EM Algorithm and Extensions, John Wiley & Sons.
- 2) Lee,B.J., Fujiwara,A. and Moon,N.G. (2001): Sensitive Analysis of Route Choice Model Using Fuzzy Reasoning Methods, 9th WCTR, Seoul.
- 3) 秋山孝正 (1997): ファジィ推論を用いた経路選択行動分析に関する研究, 科学研究費補助金基盤研究成果報告書.